

A Study of Subseasonal Predictability

MATTHEW NEWMAN, PRASHANT D. SARDESHMUKH, CHRISTOPHER R. WINKLER, AND
JEFFREY S. WHITAKER

NOAA-CIRES Climate Diagnostics Center, University of Colorado, Boulder, Colorado

(submitted to *Monthly Weather Review* June 2002; revised January 2003)

ABSTRACT

The predictability of weekly averaged circulation anomalies in the Northern Hemisphere, and diabatic heating anomalies in the Tropics, is investigated in a linear inverse model (LIM) derived from their observed simultaneous and time-lag correlation statistics. In both winter and summer, the model's forecast skill at Week 2 (Days 8 to 14) and Week 3 (Days 15 to 21) is comparable to that of a comprehensive global medium range forecast (MRF) model developed at the National Centers for Environmental Prediction (NCEP). Its skill at Week 3 is actually higher on average, partly due to its better ability to forecast tropical heating variations and their influence on the extratropical circulation. The geographical and temporal variations of forecast skill are also similar in the two models. This makes the much simpler LIM an attractive tool for assessing and diagnosing atmospheric predictability at these forecast ranges.

The LIM assumes that the dynamics of weekly averages are linear, asymptotically stable, and stochastically forced. In a forecasting context, the predictable signal is associated with the deterministic linear dynamics, and the forecast error with the unpredictable stochastic noise. In a low-order linear model of a high-order chaotic system, this stochastic noise represents the effects of both chaotic nonlinear interactions and unresolved initial components on the evolution of the resolved components. Its statistics are assumed here to be state-independent.

An average signal to noise ratio is estimated at each grid point on the hemisphere, and is then used to estimate the potential predictability of weekly variations at the point. In general, this predictability is on the order of 50% higher in winter than summer over the Pacific and North America sectors; the situation is reversed over Eurasia and North Africa. Skill in predicting tropical heating variations is important for realizing this potential skill. The actual LIM forecast skill has a similar geographical structure but weaker magnitude than the potential skill.

In this framework, predictable variations of forecast skill from case to case are associated with predictable variations of signal rather than of noise. This contrasts with the traditional emphasis in studies of shorter-term predictability on flow-dependent instabilities, i.e. on the predictable variations of noise. In the LIM, the predictable variations of signal are associated with variations of the initial state projection on the growing singular vectors of the LIM's propagator, which have relatively large amplitude in the tropics. At times of strong projection on such structures, the signal to noise ratio is relatively high, and the Northern Hemispheric circulation is not only potentially but also actually more predictable than at other times.

1. Introduction

Weather prediction is in essence a problem of predicting the conditional probability distribution P_f of future atmospheric states given the occurrence of an initial state. Predictability – defined most broadly as the potential to make a useful forecast – depends upon how different P_f is from the unconditional (i.e. climatological) distribution P (e.g., Schneider and Griffies 1999; Klemm 2002). Starting as a narrow distribution, P_f generally drifts towards P and broadens with forecast lead time due to the chaos in the atmosphere, and eventually

becomes indistinguishable from P . One may view the forecast as being “deterministic” up to such lead time as P_f is narrow enough for its width to be unimportant, and “probabilistic” thereafter. So long as it remains different from P , however, one may say that there is some predictability in the system.

The main task of any probabilistic or ensemble prediction system then is to predict P_f as accurately as possible: not only its mean and standard deviation but also higher moments, since one cannot gauge beforehand what aspect of it a user might find most useful. Attention has traditionally been given to estimating the mean and standard deviation from an ensemble of forecasts made using slightly different initial conditions from that of a control forecast. The ensemble mean can also be issued as a single “deterministic” forecast. But because nature may select any state at random from P_f , such an ensemble mean forecast from even a perfect model has limited

Corresponding author address: Matthew Newman, NOAA-CIRES Climate Diagnostics Center, Mail Code R/CDC, 325 Broadway, Boulder CO 80305-3328.
E-Mail: matt@cdc.noaa.gov

skill. Indeed one can show quite generally, for any forecast variable at any grid point, and regardless of whether or not P_f is Gaussian, that the expected correlation ρ_∞ of such ensemble-mean forecasts with observations is given by $\rho_\infty^2 = S^2/[1 + S^2]$, where $S = s/\sigma_f$ is the forecast signal to noise ratio. The signal s here is the mean of P_f , estimated as the ensemble mean, and the noise σ_f is its standard deviation, estimated as the ensemble spread. S generally decreases with forecast lead time due to both the decrease of signal and increase of noise. Note that ρ_∞ is the expected skill of a perfect model in which the signal is determined as the mean of an infinite-member ensemble. The expected skill is lower if an insufficient number of ensemble members is used to estimate the signal. The *actual* skill ρ can be brought down even further by initial condition errors and model errors. A more detailed discussion of these issues can be found in Sardeshmukh et al (2000; their sections 9 and 10).

Note that if σ_f is underestimated (as is often the case with operational systems) then S , and therefore ρ_∞ , are overestimated even when s is correctly determined. Accurate representations of both s and σ_f are thus important for estimating predictability. Accurate representation of σ_f is also directly important for correctly representing the tails of P_f , and therefore the risks of extreme anomalies.

How one quantifies predictability in a probabilistic forecasting framework depends upon the forecast problem of interest. The traditional emphasis in studies of shorter-term predictability on flow-dependent instabilities, i.e. on the predictable variations of noise, has led to a focus on measures comparing forecast and climatological variance (e.g., Schneider and Griffies 1999). Since usually forecast skill also results from a change in the mean of P_f (relative to the zero mean of P), a measure combining both signal and noise effects may be desirable, such as relative entropy R (Kleeman 2002). If the width of P_f is independent of the mean of P_f , however, then *variation* of predictability is due to variation of the signal alone. In this case, it is simpler to consider ρ_∞ , which has the additional advantage of being directly comparable to a commonly-used measure of forecast skill¹. This has also been the approach used for seasonal predictability, assessed in terms of the ratio of SST-forced variability to state-independent “internal” variability (e.g., Madden 1976; Shukla and Gutzler 1983; Chervin 1986; Zwiers 1987; Kumar et al 1996; Brankovic and Palmer 1997; Rowell 1998; Chen and van den Dool 1999; Anderson et al. 1999; Sardeshmukh et al 2000).

The particular predictability problem we will consider is that of weekly average anomalies of the northern hemispheric circulation and tropical diabatic heating at lead times of two weeks and greater, during both winter and summer. The inclusion of tropical heating is motivated by its potentially important impact upon subsea-

sonal extratropical forecast skill (e.g., Sardeshmukh and Hoskins 1988; Ferranti et al 1990; Qin and Robinson 1995; Hendon et al 2000; Winkler et al. 2001). With the exception of a few other analyses (e.g., van den Dool and Saha 1990; Branstator et al 1993; Corti and Palmer 1997), most prediction efforts and predictability studies have traditionally focused on shorter (synoptic) and longer (seasonal to interannual) time scales. Yet there is much to be said for studying predictability on the sub-seasonal scale, if only because variability on this scale accounts for a large fraction of the total tropospheric circulation variability. Also, episodes of subseasonal duration, such as springtime floods, summertime heat waves, and prolonged wet or dry spells, are phenomena with obvious societal consequences.

We will present estimates of ρ_∞ for subseasonal variability using the simple linear model presented in Winkler, Newman and Sardeshmukh (2001; hereafter WNS), and motivated further in section 2. WNS constructed a 37-component Linear Inverse Model (LIM) of wintertime weekly averaged circulation anomalies in the northern hemisphere and column-averaged diabatic heating anomalies in the tropics. The LIM technique is briefly reviewed in section 2, and model details as well as a similar LIM of summertime weekly averaged anomalies are presented in section 3. Section 4 provides further evidence that the skill of these simple LIMs in predicting Week 2 and Week 3 anomalies is comparable to that of the NCEP medium range forecast (MRF) model. Predictability ρ_∞ estimated in the manner outlined above is shown in section 5, and compared with actual average (cross-validated) LIM skill ρ . It is shown that temporal and geographical variation of ρ_∞ can explain similar variation of ρ , and that much of this variation is in turn due to variations in tropical heating. It is finally shown that most of the predictability is associated with three growing singular vectors of the LIM’s propagator matrix. When the initial projection on these singular vectors is high, not only the potential but also the actual forecast skill is considerably higher than at other times. That is, both ρ_∞ and ρ can on occasion be substantially larger than their average values, and such occasions can be identified *a priori*.

2. Motivation

The general broadening and drift of P_f towards P associated with loss of predictability are symptomatic of chaos in the atmosphere. But they are also characteristic of any *stochastically forced dissipative process*, whose expected mean, given an initial condition, drifts (not necessarily monotonically) towards climatology and whose variance about this mean increases in response to the stochastic forcing, eventually saturating at its climatological value. A linear one-dimensional autoregressive process, also known as an Ornstein-Uhlenbeck process or simply as red noise, is the simplest example of this. In approximating atmospheric variability as such a pro-

1. Also, note that for the univariate case, if $S(\tau) \ll 1$ and σ_f approaches climatological variance, then $\rho_\infty \approx 2R$.

cess, the dissipative tendency (i.e the drift towards climatology) is associated with thermal and mechanical damping, and the stochastic forcing with rapidly varying chaotic fluxes of heat, momentum, and moisture. One may imagine how a stochastically forced dissipative model might thus capture the essence of the loss of predictability with lead time.

An obvious limitation of a one-dimensional red-noise model in this regard is that its expected mean does drift monotonically towards climatology. One can, however, readily construct more complicated but still very simple *multidimensional* stochastically forced linear dissipative models in which S can decrease non-monotonically, and even exhibit quasi-oscillatory behavior, before eventually becoming small. Furthermore, in many systems the effective linear operator governing this process, as well as the statistics of the stochastic forcing, may be estimated directly from the observed simultaneous and time-lag covariance statistics of the system variables, as described for example in Penland and Sardeshmukh (1995). An effectively linear, stochastically forced model of a system thus constructed is called a Linear Inverse Model (LIM) (see also Penland 1989; Penland and Ghil 1993; Penland 1996; Delsole and Hou 1999; and WNS).

In any multidimensional statistically stationary system with components x_i , one may define a time-lag covariance matrix $\mathbf{C}(\tau)$ with elements $C_{ij} = \langle x_i(t+\tau) x_j(t) \rangle$, where angle brackets denote a long-term average. In linear inverse modeling, one assumes that the system satisfies $\mathbf{C}(\tau) = \mathbf{G}(\tau)\mathbf{C}(0)$, where $\mathbf{G}(\tau) = \exp(\mathbf{B}\tau)$ and \mathbf{B} is a constant matrix, and uses this relationship to estimate \mathbf{B} given observational estimates of $\mathbf{C}(0)$ and $\mathbf{C}(\tau_0)$ at some lag τ_0 . In such a system any two states separated by a time interval τ are related as $\mathbf{x}(t+\tau) = \mathbf{G}(\tau) \mathbf{x}(t) + \boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon}$ is a random error vector with covariance $\mathbf{E}(\tau) = \mathbf{C}(0) - \mathbf{G}(\tau)\mathbf{C}(0)\mathbf{G}^T(\tau)$. Note that the system need not have Gaussian statistics for these relations to hold. However, for its statistics to be stationary, \mathbf{B} must be dissipative, i.e its eigenvalues must have negative real parts. In the context of our preceding discussion, we may write the unconditional and conditional probability distributions for $\mathbf{x}(t+\tau)$ given $\mathbf{x}(t)$ as $P\{0, \mathbf{C}(0), \dots\}$ and $P_f\{\mathbf{G}(\tau)\mathbf{x}(t), \mathbf{E}(\tau), \dots\}$ respectively. Thus $\mathbf{G}(\tau)\mathbf{x}(t)$ represents an “infinite-member ensemble-mean” forecast for $\mathbf{x}(t+\tau)$ given $\mathbf{x}(t)$, i.e the forecast signal, and $\mathbf{E}(\tau)$ represents the expected covariance of its error, i.e the noise. Note that for large lead times τ , $\mathbf{G}(\tau)\mathbf{x}(t) \rightarrow 0$ and $\mathbf{E}(\tau) \rightarrow \mathbf{C}(0)$, so $P_f \rightarrow P$ as required.

The underlying dynamical equation of this system is

$$\frac{d\mathbf{x}}{dt} = \mathbf{B}\mathbf{x} + \mathbf{F}_s \quad (1)$$

where \mathbf{F}_s is noise which is white in time but which may be spatially correlated, and related to the forecast error vector as

$$\boldsymbol{\varepsilon} = \mathbf{G}(\tau) \int_t^{t+\tau} \mathbf{G}(t')^{-1} \mathbf{F}_s(t') dt'. \quad (2)$$

The theory which allows a highly nonlinear system to be represented by (1) (e.g., Papanicolaou and Kohler 1974; Hasselmann 1976; see also Penland 1996) requires \mathbf{x} to be “coarse-grained”; that is, \mathbf{x} represents spatial and/or temporal averages. The noise then results from the effects of initial uncertainty, not only because there is an unresolved remainder to \mathbf{x} but also because information regarding the precise state is lost in the average. LIM then makes the important additional assumption that \mathbf{F}_s is state-independent. This is a common although not necessary restriction to (1), but the assumption makes the empirical determination of \mathbf{B} possible with the limited amount of data available; its validity can also be checked *a posteriori*.

3. Model details and data

a) LIM

As in WNS, we define \mathbf{x} as

$$\mathbf{x} \equiv \begin{bmatrix} \Psi \\ \mathbf{H} \end{bmatrix}, \quad (3)$$

where Ψ is anomalous streamfunction and \mathbf{H} is anomalous tropical diabatic heating. Both quantities here represent 7-day running mean averages. Equation (1) is then

$$\frac{d}{dt} \begin{bmatrix} \Psi \\ \mathbf{H} \end{bmatrix} = \begin{bmatrix} \mathbf{B}_{\Psi\Psi} & \mathbf{B}_{\Psi\mathbf{H}} \\ \mathbf{B}_{\mathbf{H}\Psi} & \mathbf{B}_{\mathbf{H}\mathbf{H}} \end{bmatrix} \begin{bmatrix} \Psi \\ \mathbf{H} \end{bmatrix} + \begin{bmatrix} \mathbf{F}_{s\Psi} \\ \mathbf{F}_{s\mathbf{H}} \end{bmatrix} \quad (4)$$

and infinite-member ensemble-mean forecasts are made as

$$\widehat{\mathbf{x}}(t+\tau) = [\mathbf{G}(\tau_0)]^{\tau/\tau_0} \mathbf{x}(t) = \exp\left(\begin{bmatrix} \mathbf{B}_{\Psi\Psi} & \mathbf{B}_{\Psi\mathbf{H}} \\ \mathbf{B}_{\mathbf{H}\Psi} & \mathbf{B}_{\mathbf{H}\mathbf{H}} \end{bmatrix} \tau\right). \quad (5)$$

Here $\mathbf{F}_{s\Psi}$ and $\mathbf{F}_{s\mathbf{H}}$ are the white noise forcing of Ψ and \mathbf{H} , respectively. Note that including \mathbf{H} explicitly in \mathbf{x} makes it possible to diagnose how tropical heating impacts streamfunction predictability through $\mathbf{B}_{\Psi\mathbf{H}}$ (Newman et al. 2000; WNS).

Certainly, other variables not included in (3) could also be important to the evolution of Ψ or \mathbf{H} . The inverse model does, however, implicitly include the effects of all other variables linearly related to Ψ and/or \mathbf{H} . This is an important distinction from a forward dynamical model in which the evolution of the state vector is governed only by the *explicitly* represented interactions among its

components.

Linear inverse models for winter (December 1–February 28) and summer (June 1–August 31) have been constructed using global National Centers for Environmental Prediction (NCEP) data from 1969–2000. [Unlike in WNS, the reanalysis data used had the TOVS data assimilation error fixed.] Diabatic heating rates were diagnosed from an improved iterative solution of the “chi-problem” (Sardeshmukh 1993; Sardeshmukh et al. 1999), as described in WNS. Low-frequency anomalies were defined by removing the seasonal cycle (that is, the annual mean plus the first three annual harmonics) from each variable at each gridpoint and then applying a 7-day running mean filter. Streamfunction anomalies were determined at 750 hPa and 250 hPa, and then spectrally smoothed to T21 resolution. Column-integrated diabatic heating anomalies from the surface to the tropopause, a simple measure of tropical forcing which provides a single measure of \mathbf{H} in (3) at each horizontal grid point, were spatially smoothed with a spectral filter which attenuates small-scale features and Gibbs phenomena (Sardeshmukh and Hoskins 1984).

Anomalies were projected onto their leading Empirical Orthogonal Functions (EOFs), determined separately for winter and summer data. The domain for streamfunction EOFs was the Northern Hemisphere, and for diabatic heating EOFs the region 30°S–30°N. The leading 30 EOFs of anomalous streamfunction were retained in each model. For anomalous tropical diabatic heating, the leading 7 (20) EOFs of anomalous tropical diabatic heating were retained in winter (summer). The time-varying coefficients of these EOFs, i.e., the principal components (PCs), define a 37 (50)-component state vector \mathbf{x} in winter (summer). The retained streamfunction EOFs account for more than 90% of the low-frequency variability in regions of large streamfunction variability. In winter, the 7 heating EOFs capture over 70% of the low-frequency variability in the central and western Tropical Pacific, though only about 35% of the domain-integrated variability. In summer, the 20 heating EOFs capture almost 60% of the domain-integrated variability. The separate truncation of these models was dictated mainly by the desire to produce the best streamfunction forecasts over the hemisphere according to the skill measures and sampling constraints described in WNS. The retention of 13 additional EOFs in summer made only a small (generally less than 5%) difference in these skill measures. Qualitatively, our results are fairly insensitive to the EOF truncation of both ψ and \mathbf{H} . Tests of robustness of the LIM, including quantitative sensitivity of results to EOF truncation, τ_0 , and the heating dataset used, as well as the cross validation technique¹, are discussed at length in WNS.

1. The cross-validation technique used in WNS was modified for this paper so that separate EOF bases and 29-year climatologies were used for each LIM. The resulting difference in cross-validated forecast skill was exceedingly minor (on the order of 1–2%) except for January–February 1999.

TABLE 1. MRF98 model details.

Horizontal Resolution:	Triangular truncation at total wavenumber 62
Vertical Resolution	28 sigma levels
Convective Parameterization	Simplified Arakawa-Schubert (Pan and Wu, 1994) for deep convection, vertical diffusion of heat and moisture in the upper boundary layer and lower free atmosphere (Tiedtke 1983) to simulate non-precipitating shallow convection.
Radiation	Short-wave due to Lacis and Hansen (1974), Long-wave due to Fels and Schwarzkopf (1975).
Boundary Layer Parameterization	Based on Troen and Mahrt (1996).
Topography	mean orography
Cloud scheme	3 levels of stratiform clouds, fractional coverage determined as a quadratic function of layer RH.
Time integration	semi-implicit, taking into account physics tendencies
Horizontal diffusion	Laplacian type, based on Leith (1971) formulation.

b) GCM

Forecasts made using different versions of NCEP’s nonlinear Medium-Range Forecast (MRF) general circulation model (GCM) were also analyzed. Week 2 forecast skill of 11-member ensemble mean forecasts of the MRF model used operationally at NCEP, for the winters of 1996/97–1999/00 and the summers of 1996–1999, was determined. Since operational MRF forecasts were not available beyond week 2, and only forecasts starting in 1996 were available, a parallel analysis using a version of the MRF model that was operational at NCEP during January–June 1998 (Caplan et al 1997, Wu et al 1998 and references therein) was conducted. An overview of the model formulation is presented in Table 1. This version of the MRF (hereafter “MRF98”) was used to make 21-day forecasts initialized with the NCEP/NCAR Reanalysis (Kalnay et al 1996) once each day at 00 UTC. More recent versions of the MRF are not compatible with the initial condition files available with the reanalysis. Integrations have been run from November 1978 to the present. Ultimately, this “reforecast” dataset will consist of an ensemble of forecasts, but the results presented here are for the (single) control run. [This dataset of forecasts using a frozen model is publicly available via the Climate Diagnostics Center (CDC) web site (<http://www.cdc.noaa.gov/~jsw/refcst>), and can be used for predictability studies as well as for calibrating and correcting real-time forecasts with the

same model.]

As do all GCMs, the MRF has a climate drift which increases with forecast lead. The 20+ yr period of the MRF98 forecasts was used to determine this evolving model climate at each grid point as the seasonal cycle of all forecasts for the years 1979-2000. Since this systematic error was also a function of forecast lead, it was computed separately for all Week 2 and Week 3 forecasts. All MRF98 forecasts were then corrected by subtracting this error from each forecast at each grid point. A similar correction was not attempted for the operational MRF forecasts, due to their limited period and unfixed forecast system.

The different forecast models used in this paper are summarized in Table 2.

4. Forecast skill

Our measure of forecast skill is the average anomaly correlation between $\widehat{\mathbf{x}}(t + \tau)$, the ensemble mean forecast at lead τ , and $\mathbf{x}(t + \tau)$, the observed state at that time:

$$\rho(\tau) = \frac{\langle \widehat{\mathbf{x}}(t + \tau), \mathbf{x}(t + \tau) \rangle}{[\langle \widehat{\mathbf{x}}(t + \tau), \widehat{\mathbf{x}}(t + \tau) \rangle \langle \mathbf{x}(t + \tau), \mathbf{x}(t + \tau) \rangle]^{1/2}} \quad (6)$$

where $\langle \bullet, \bullet \rangle$ is an inner product which can be chosen for any norm of interest. Spatial dependence of skill is measured with local anomaly correlation, with $\langle \bullet, \bullet \rangle$ defined as a scalar product of the forecast and verification anomaly time series computed at each grid point. Similarly, temporal dependence is measured with anomaly pattern correlation, with $\langle \bullet, \bullet \rangle$ defined at each time as a scalar product of the forecast and verification anomalies in the region north of the Equator between 120E and 60W. Note that all measures of actual forecast skill presented in this study for the LIM are based upon cross-validated (jack-knifed) forecasts. Also, forecasts are always compared with the complete (that is, the untruncated) observations in grid space.

By both measures, LIM forecasts of streamfunction at any lead τ are more skillful than persistence, climatology, and the local time-lag autocorrelation. They also show much higher skill than simplified linear dynamical models, including those forced with observed tropical heating (WNS).

To use this stochastically-forced linear model to

TABLE 2. Summary of forecast models used in this paper.

Model	Forecast lead (weeks)	Ensemble size	Period
LIM	2,3	∞	12/1969-2/2000
MRF98	2,3	1	12/1978-2/2000
OMRF	2	11	12/1996-2/2000

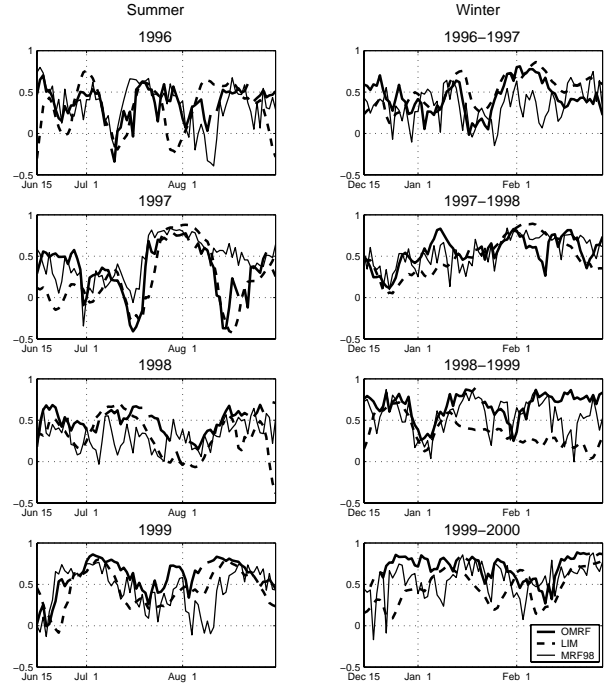


FIG. 1: Spatial pattern correlation of week 2 250 hPa streamfunction forecasts in the region 0–90N, 120E–60W from the linear inverse model (LIM), the operational MRF ensemble mean, when available (OMRF), and the single MRF reforecast run (MRF98). Verifications begin December 15 of each year for winter and June 15 for summer.

investigate predictability, its forecast skill must be competitive with that of fully nonlinear GCMs. This is indeed the case. Shown in Fig. 1 is Week 2 forecast skill from LIM and from 11-member ensemble mean operational MRF forecasts, for the winters of 1996/97-1999/00 and the summers of 1996-1999. For GCM forecasts, we define week 2 as the mean of days 8–14 and week 3 as the mean of days 15–21, where day 0 is the forecast initialization. The LIM forecast is initialized on day -3 with the 7-day mean centered on that date (that is, the mean of days -6–0), and is then run forward 14 and 21 days. This definition is illustrated in Fig. 2. Thus, the LIM and GCM forecasts verify for the same period, and neither uses information past day 0 in the initialization.

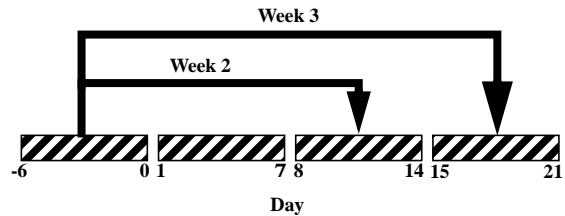


FIG. 2: Schematic illustrating forecast protocol used in this paper. Note that the MRF is initialized on Day 0.

Anomaly correlation of forecasts

Based on forecasts made during DJF 1978/79-1999/2000

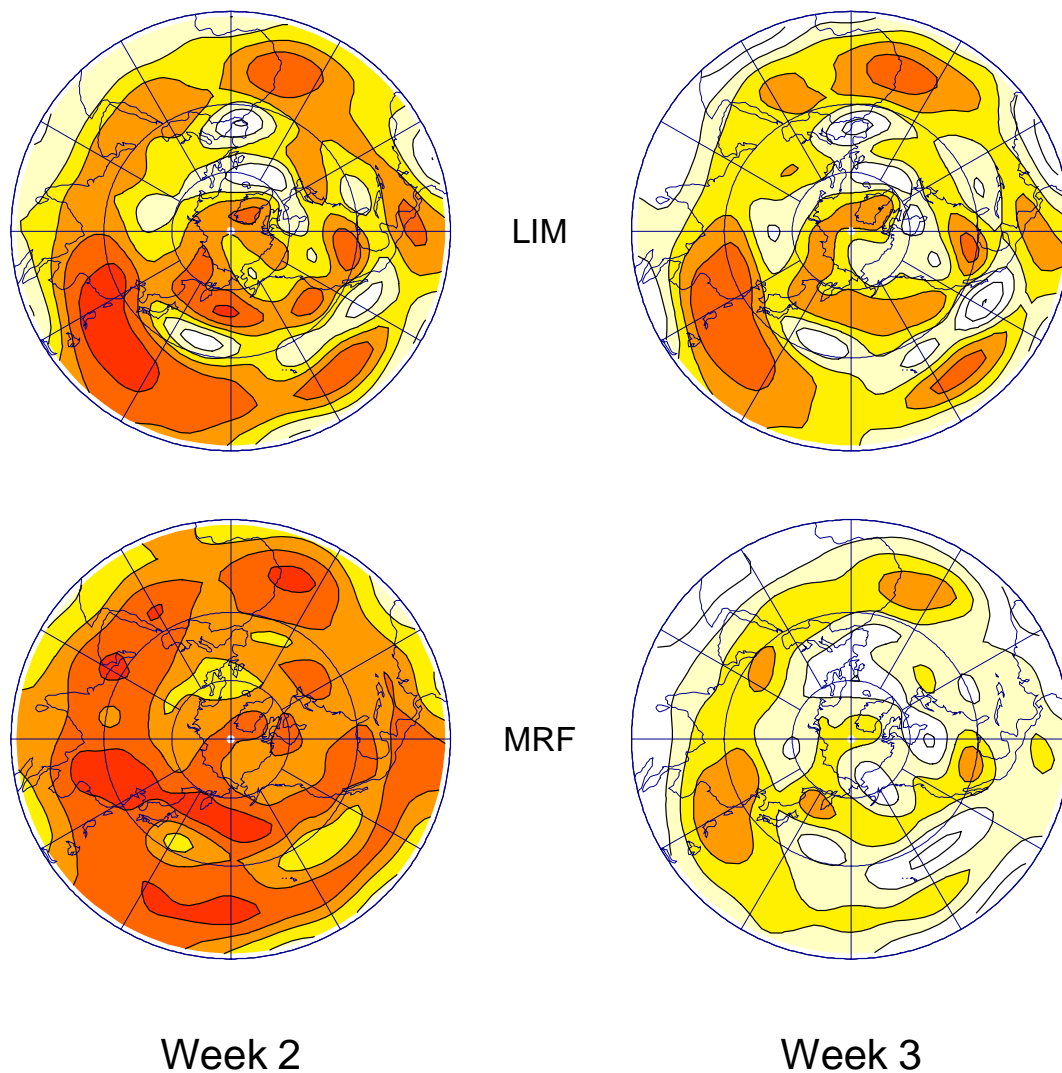


FIG. 3a: Comparison of local anomaly correlation of 250 hPa streamfunction wintertime forecasts for the LIM and MRF98. Top: LIM. Bottom: MRF98. Contour interval is 0.1 with negative and zero contours indicated by blue shading and dashed lines. Shading of positive values starts at 0.2; redder shading denotes larger values of correlation, with the reddest shading indicating values above 0.6.

As a result, the week 2 predictions begin verifying on December 15 each winter and June 15 each summer.

Skill of the two models is comparable. Occasionally, LIM forecasts have higher skill than MRF ensemble-mean forecasts, although there are clearly times when the LIM is much worse than the MRF. It is particularly striking that there are extended periods during which both models have similar variations in skill, raising the possibility that they may be nearing the limit of predictability at these times. It should nonetheless be noted that, since the two models share similar streamfunction initial conditions, similar forecast skill may also be partly due to systematic error in those initial conditions.

How the LIM compares with the MRF ensemble at longer forecast leads is not known since operational MRF forecasts are not available beyond week 2. Also, sampling might impact the comparison between the two models since only four years of forecasts were used. Thus, the LIM is next compared with the MRF98 forecasts. Although currently only a single 21-day MRF98 forecast made from each day is available, the skill of these “single-member ensemble forecasts” at week 2 was often essentially identical to the operational ensemble (see Fig. 1). Averaged over these four years (and keeping in mind that the operational forecast system was

Anomaly correlation of forecasts

Based on forecasts made during JJA 1979-2000

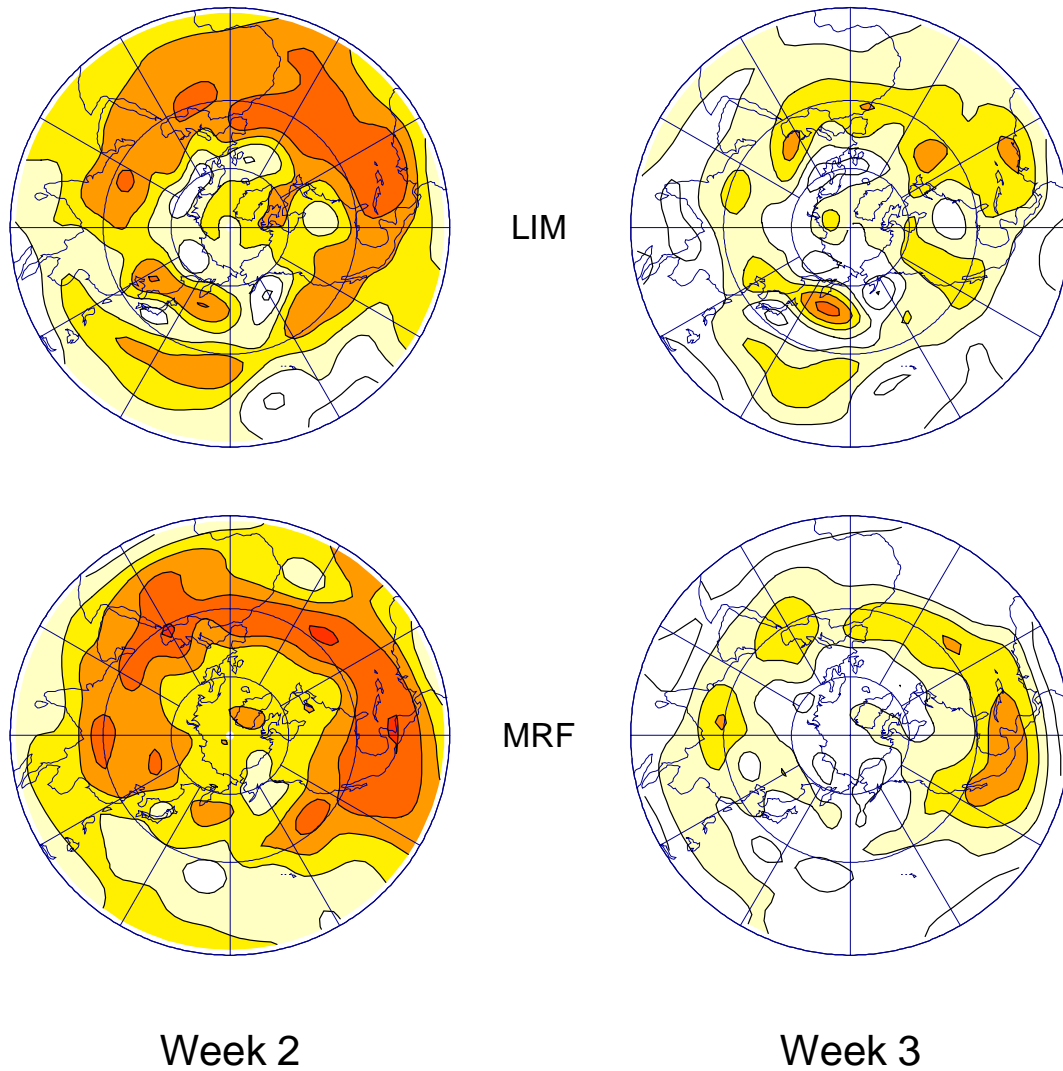


FIG. 3b: Same as Fig. 3a but for the summertime.

not fixed during that time), the week 2 skill of the operational ensemble mean was about 17% higher than that of the reforecast.

The LIM and MRF98 forecasts again have comparable skill. Particularly striking is that the geographic distribution of skill (shown in Fig. 3), as well as its seasonality, is so similar between these two very different models. For example, during winter both models have a banded structure in skill across the Pacific, with a minimum in a band along 30°N , particularly on either side of the ocean basin. Maxima over western Africa and the North Atlantic, and a minimum over Northern Europe, are also evident. During summer, both models have greatest skill in a zone extending from Central

America eastwards to North Africa and in a region over the northwest Pacific, with a pronounced minimum just south of Japan.

The seasonal means of pattern correlation for the LIM, MRF98, and (when available) MRF operational ensemble for each year are plotted in Fig. 4. The averages of ρ over all years, and over the years of the MRF98 runs, are listed in Table 3 for winter and Table 4 for summer. The substantial interannual variability of forecast skill in both models is quite similar, particularly in winter; the correlation between seasonally-averaged LIM and MRF98 wintertime week 2 (3) skill is 0.73 (0.79).

LIM forecasts of similar skill to the MRF98 may

TABLE 3. Mean pattern correlation between forecasts and verification, in the Northern Hemisphere between 120E and 60W during winter, compared with the expected skill determined from the LIM.

	Week 2	Week 3
ρ_{∞} (LIM), 1969/70-1999/00	0.55	0.45
ρ (LIM), 1969/70-1999/00	0.39	0.33
ρ_{∞} (LIM), 1978/79-1999/00	0.52	0.44
ρ (LIM), 1978/79-1999/00	0.37	0.30
ρ (MRF), 1978/79-1999/00	0.48	0.24

result when much of the chaotic nonlinear development of weekly-averaged anomalies can be parameterized effectively as linear terms plus white noise. LIM forecasts better than MRF98 forecasts at week 3 could, however, be due to deficiencies in the GCM itself. Poor MRF98 forecasts of tropical heating, particularly of the MJO, is one possibility (e.g., Sardeshmukh and Hoskins 1988; Ferranti et al. 1990; Klinker 1990; Cai et al. 1996; Hendon et al. 2000), as is shown in Fig. 5. MRF98 forecasts of tropical heating are considerably worse than those of the LIM over the west Pacific where the MRF98 has poor skill as early as week 1 (not shown) and no skill by week 2. [Tropical precipitation forecasts from the operational MRF ensemble are similarly poor

TABLE 4. Same as Table 3 except for summer.

	Week 2	Week 3
ρ_{∞} (LIM), 1969-1999	0.47	0.36
ρ (LIM), 1969-1999	0.29	0.21
ρ_{∞} (LIM), 1979-1999	0.45	0.35
ρ (LIM), 1979-1999	0.29	0.20
ρ (MRF), 1979-1999	0.34	0.17

(not shown).] During summer, the MRF98 has somewhat better skill in the Tropics, although still poorer than the LIM.

Systematic error may also reduce MRF98 skill. For example, in summer forecast errors in the Pacific jet are very large, as can be seen in 21-yr averages of week 2 forecasts of relative vorticity during February and June (Fig. 6). This impacts both the scale and propagation of Rossby waves across the Pacific waveguide (Newman and Sardeshmukh 1998), resulting in errors not correctable by removal of the systematic error alone. In contrast, the LIM has no mean error, by design.

The MRF98 should have higher skill than shown here if each reforecast were an ensemble rather than a single realization. However, some regional and temporal differences in week 3 skill between the LIM and the MRF98 are too large to be completely eliminated by this possibility (Sardeshmukh et al. 2000), as is the sharp decline in MRF98 skill from week 2 to week 3 especially when compared to the much smaller decline for the LIM. Furthermore, the relatively small improvement

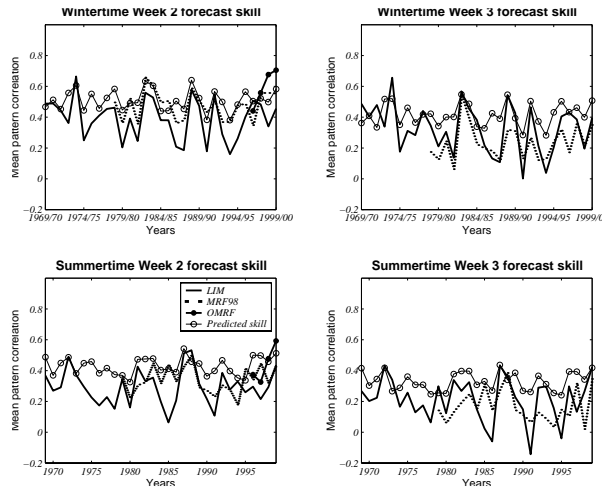


FIG. 4: A comparison of the LIM and MRF forecast skill computed separately for each year of the jack-knifed forecasts. The solid line indicates the seasonal mean of daily spatial pattern correlation of week 2 and week 3 250 hPa streamfunction LIM forecasts in the region 0–90N, 120E–60W. The same forecast quantity is indicated for the MRF reforecasts (MRF98) by the dotted line, and for the operational MRF ensemble mean (OMRF; at week 2 only) by the solid line with filled circles. Also shown is the theoretical forecast skill, indicated by the seasonal mean of ρ_{∞} , for the same forecast measure, as the thin solid line with open circles.

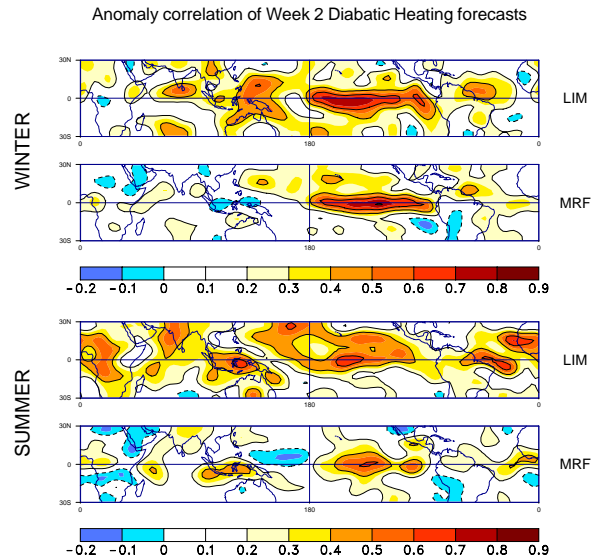


FIG. 5: Comparison of local anomaly correlation of column-integrated tropical diabatic heating week 2 forecasts for the LIM and MRF98. Top: wintertime. Bottom: summertime. Contour (fill) interval is 0.2 (0.1).

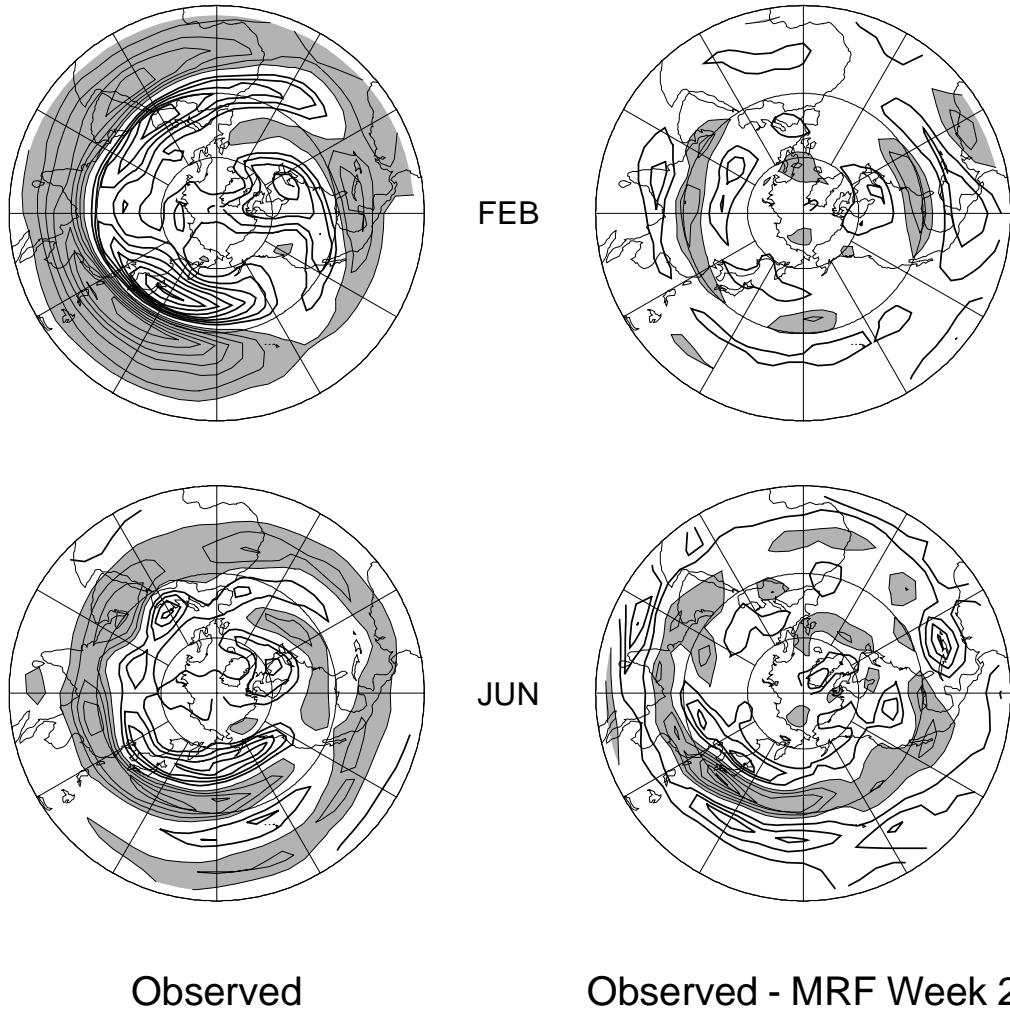


FIG. 6: Comparison of observed monthly mean 250 hPa relative vorticity during February and June with that predicted at week 2 by the MRF98. Left panels show observed vorticity; contour interval is $6 \times 10^{-6} \text{ s}^{-2}$. Right panels show difference maps between the observed vorticity and the MRF week 2 vorticity, determined as the average of all week 2 forecasts verifying during the month; contour interval is $3 \times 10^{-6} \text{ s}^{-2}$. In all four panels, negative values are indicated by shading and thinner contours; the zero contour is eliminated for clarity

in skill actually obtained on average by the operational ensemble relative to the MRF98 reforecast at week 2 suggests that at least some of the higher LIM week 3 skill is real.

5. Predictability in the linear system

Clearly, the LIM does not reproduce every facet of the MRF98 forecast skill, but Figs. 3 and 4 show that much of the temporal and spatial distribution of skill is similar between the two models. This, combined with the LIM's generally higher level of skill in the extended range, minimal systematic error, appropriate forecast spread, and better simulation of tropical heating, suggests that

further exploration of the LIM predictability limits and the sources of that predictability can give significant insight into the potential skill of subseasonal forecasts.

Recall that we will measure predictability by the expected skill of a perfect model infinite-member forecast ensemble,

$$\rho_{\infty}(\tau) = \frac{S^2(\tau)}{\{[S^2(\tau) + 1]S^2(\tau)\}^{1/2}} \quad (7)$$

where $S(\tau)$ is the signal-to-noise ratio and τ is the forecast lead. We determine $S(\tau)$ either for each system component (e.g., at each gridpoint for each variable) or for the system as a whole over some geographical

Anomaly correlation of Week 3 forecasts

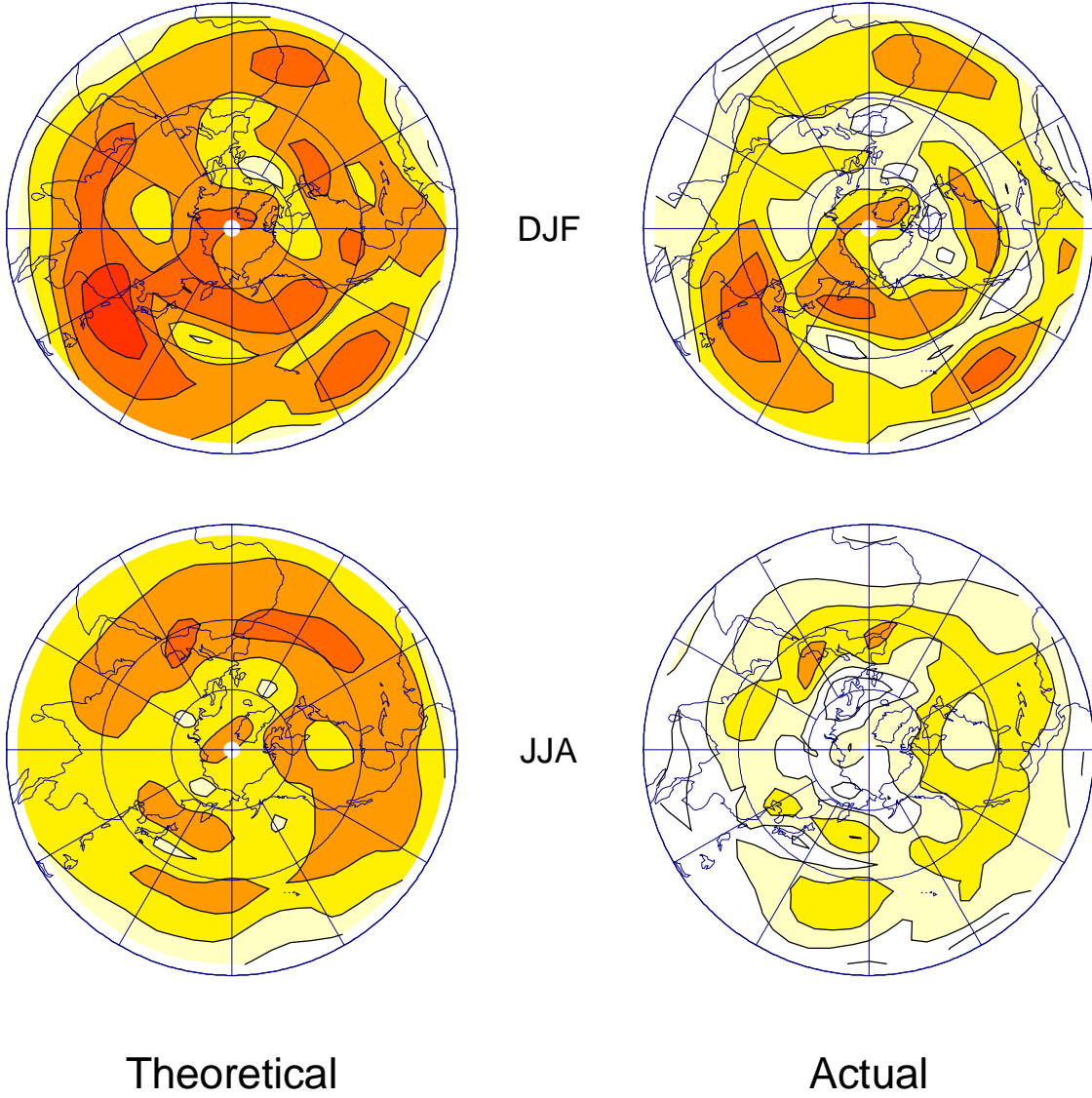


FIG. 7: Local anomaly correlation of week 3 streamfunction forecasts for LIM, for winter (top) and summer (bottom), at 250 hPa. Left: Theoretical mean predictability limit ρ_∞ . Right: Actual skill from 30 years of jack-knifed forecasts. Plotting conventions are as in Fig. 3.

region. If $\rho_\infty(\tau)$ represents local anomaly correlation, $S(\tau)$ is a function of space but not time:

$$S^2(\tau) = \frac{F(\tau)_{ii}}{E(\tau)_{ii}}, \quad (8)$$

where $\mathbf{F}(\tau) = \langle \widehat{\mathbf{x}}(t+\tau) \widehat{\mathbf{x}}(t+\tau)^T \rangle$ is the signal covariance matrix, $\mathbf{E}(\tau) = \mathbf{C}(0) - \mathbf{G}(\tau)\mathbf{C}(0)\mathbf{G}^T(\tau)$ is the covariance matrix of the forecast error, and i is the grid index. That is, S is a ratio of the diagonal elements of \mathbf{F} and \mathbf{E} . If instead $\rho_\infty(\tau)$ represents pattern correlation, $S(\tau)$ is a function of time but not space:

$$S^2(\tau) = \frac{\text{tr}[\mathbf{F}(\tau)]}{\text{tr}[\mathbf{E}(\tau)]} \quad (9)$$

where the trace again is defined only between 120 E and 60 W, poleward of the equator. That is, S is a ratio of the sum of the diagonal elements of \mathbf{F} and \mathbf{E} . Note that for both (8) and (9), the denominators are assumed constant and are determined from statistics of the entire 30 yr period, consistent with the assumption of state-independent noise.

Since (7) is derivable from (6), dependent upon nei-

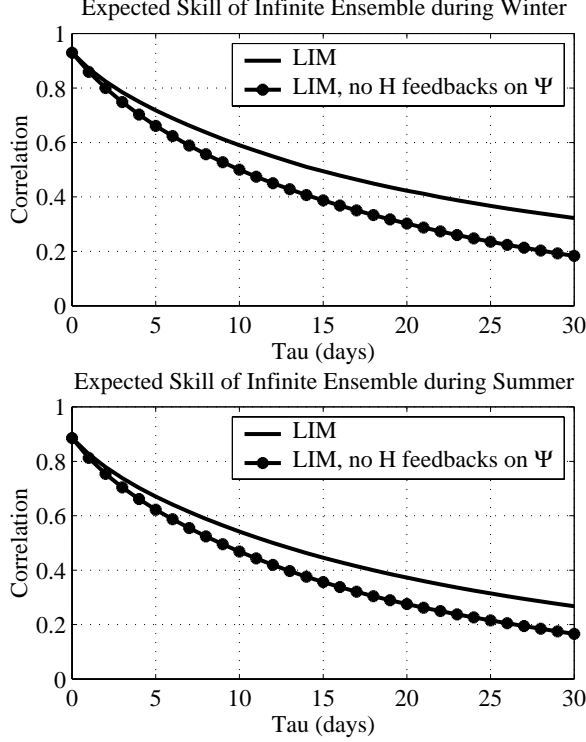


FIG. 8. Expected forecast skill as a function of lead time (days) for the LIM and for the LIM with tropical forcing removed. See text for further details.

ther linearity nor Gaussianity (Rowell 1998; Sardeshmukh et al. 2000), we will relate variations in actual (imperfect model) LIM forecast skill $\rho(\tau)$ with variations in expected (perfect model) skill $\rho_\infty(\tau)$.

5.1. Signal-to-noise ratio and expected predictability

Overall mean expected skill is determined by averaging forecast covariance over all possible initial conditions,

$$\mathbf{F}(\tau) = \mathbf{G}(\tau) \langle \mathbf{x}(t) \mathbf{x}(t)^T \rangle \mathbf{G}^T(\tau) = \mathbf{G}(\tau) \mathbf{C}(0) \mathbf{G}^T(\tau).$$

Fig. 7 compares the map of ρ_∞ , determined using (8) for forecast lead $\tau = 21$ d, with the actual anomaly correlation ρ of all cross-validated week 3 forecasts. In both seasons ρ is generally less than ρ_∞ . Certainly, while the LIM may be a good model of variability of \mathbf{x} , it is not a perfect one. Also, practical limitations to the empirical determination of \mathbf{B} could be expected to produce errors both in model formulation and model forecasts. For both these reasons, treating the LIM as if it were a perfect model underestimates the actual forecast error variance by about 10% (see also the longer discussion by WNS), and there is a weak (0.1–0.3) correlation between the actual forecast errors and the forecasts.

Despite these limitations, the patterns of actual and expected skill are very similar, and the expected week 3 skill is at least partly realized. Subseasonal predictability limits undergo strong variation from place to place

and from winter to summer. Notably, there is an overall shift of predictability from the western hemisphere (roughly 60E–120W) in the winter to the eastern hemisphere (roughly 120W–60E) in the summer. For example, 250 mb streamfunction anomalies over China and the U.S. Great Plains are much more predictable in winter than summer, whereas the situation is reversed over Europe and North Africa. On the other hand, in both seasons predictability is low east of Japan while it is high northeast and southeast of Japan. From winter to summer, there also appears to be a northward shift of predictability over Africa and the Pacific.

5.2. Impact of Tropical Heating upon Predictability

Tropical diabatic heating is strongly related to growth of streamfunction anomalies during winter and increases both forecast skill and forecast amplitude (WNS), so we expect it to contribute to predictability. The effect of tropical heating is gauged by determining ρ_∞ separately from \mathbf{B} and from a new linear operator $\hat{\mathbf{B}}$, identical to \mathbf{B} except that $\mathbf{B}_{\Psi\mathbf{H}} = 0$ in (4). Mean expected pattern correlation skill as a function of forecast lead, with and without tropical heating effects upon streamfunction included in the LIM, is shown in Fig. 8. On average, tropical heating increases predictability for all forecast leads and thus extends the period over which useful forecasts might be made. For example, during winter the full LIM has a mean expected pattern correlation of $\rho_\infty \geq 0.5$ for forecast leads up to 15 days, which is 5 days longer than for forecasts made using $\hat{\mathbf{B}}$. The impact of tropical heating is relatively as strong during summer, but overall predictability is less.

This point is further illustrated by Fig. 9, which on the left displays the forecast lead at which ρ_∞ falls below 0.5 for both wintertime and summertime 250 hPa streamfunction anomalies. On the right is shown the contribution of tropical heating to this expected predictability; that is, the difference between the potential predictability using either \mathbf{B} or $\hat{\mathbf{B}}$ as the forecast operator. During winter, tropical heating is responsible for the substantial increase in predictability over the Western Hemisphere, creating maxima which would not otherwise exist throughout the Pacific Basin. Heating also increases predictability over western Africa. During summer, the impacts of tropical heating are more subtle. Even without heating, predictability during summer would be higher in the Eastern Hemisphere, albeit in a region shifted further to the south and east. Still, heating does notably enhance predictability over broad areas of the Pacific. Without heating, week 2 predictability would not exist over the southern half of North America.

5.3. Predictability associated with maximally amplifying structures

Given that ρ_∞ depends on the magnitude of the signal, we expect that some states will be more predictable

than the limits determined above, which were averaged over *all* possible initial states. Although \mathbf{B} is asymptotically stable, long-term (if finite-time) anomaly growth is possible via modal interference. In that case, a singular value decomposition (SVD) of the propagator $\mathbf{G}(\tau) = \exp(\mathbf{B}\tau)$ under a suitable norm will produce some singular vector pairs $\mathbf{u}_i, \mathbf{v}_i$ whose singular values Σ_i (sometimes called the finite-time Lyapunov exponents) are larger than 1. The pair associated with the largest (or leading) singular value Σ_1 describe the greatest *possible* growth in the norm over the time interval $[t, t + \tau]$ (e.g., Farrell 1988; Lacarra and Talagrand 1988; Sardeshmukh et al 1997). Results of such an analysis for the L2 norm of Northern Hemisphere streamfunction (see also WNS) with $\tau=21$ d are shown for each season in Fig. 10. The leading right singular vector \mathbf{v}_1 for τ greater than about a week is dominated by strong anomalous tropical heating (Figs. 10a and b), with relatively more heating west of the dateline during summer than winter. In both seasons, the initial streamfunction anomaly (not shown) is minimal. The global scale streamfunction anomaly that results 21 days later ($\Sigma_1 \mathbf{u}_1$; Figs. 10c and d) has a pronounced zonally symmetric component, is equivalent barotropic during winter but baroclinic (not shown) during summer, and is more prominent over the Pacific (Atlantic) sector during winter (summer).

The regions of greatest anomaly growth are not necessarily the regions of greatest predictability, since the noise also has spatial structure. Thus, local predictability associated with these maximally amplifying structures was determined from (7) and (8) by setting the initial condition $\mathbf{x}(t) = \alpha_1 \mathbf{v}_1$, where $(\alpha_1 \Sigma_1)^2 = \text{tr}[\mathbf{C}(0)]$, so that $\mathbf{F}(\tau) = \alpha_1^2 \mathbf{G}(\tau) \mathbf{v}_1 \mathbf{v}_1^T \mathbf{G}(\tau)^T$. The resulting maps of ρ_∞ (Figs. 10e and f) are indeed consistent with but not entirely coincident with the evolved anomalies (Figs. 10c and d). Clearly, much of the geographic and seasonal variation of predictability in Fig. 7, such as the shift in expected skill from the Western to Eastern hemispheres from winter to summer, is due to initial states with strong projection on \mathbf{v}_1 (cf. Fig. 10 and the left panels of Fig. 7) and hence high initial amplitude of tropical heating (cf. Fig. 10 and the right panels of Fig. 9).

Note that both winter and summer leading singular vectors have large central/east Pacific heating anomalies, characteristic of ENSO. This suggests that predictability is greatest during ENSO years. However, this conclusion is only partly true, particularly for week 2. First, as WNS stressed, streamfunction anomaly growth results directly from the heating anomaly, not the SST anomaly. That is, initial anomalies can occasionally have a large projection on the leading singular vector even during non-ENSO years, and conversely projection on the leading singular vector is not always large during ENSO years (e.g., the first half of DJF 97/98). Also, during periods of high MJO activity (for example, DJF 96/97), some phases of the MJO will have a high projection on both singular vectors 1 and 2, resulting in high predictability over periods substantially shorter than a season.

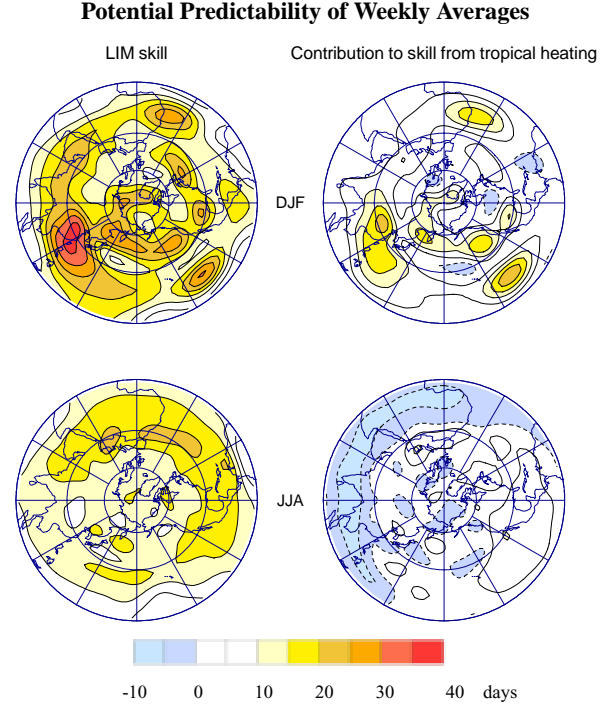


FIG. 9: Mean potential predictability limit: forecast lead at which skill (i.e., the correlation of observed and predicted anomalies) drops below 0.5. Contour interval is 5 days, with negative values indicated by blue shading and dashed lines. (a) Wintertime predictability, determined from the full LIM. (b) Change in wintertime mean predictability due to the effect of tropical heating (that is, from the $\mathbf{B}_{\text{H}\Psi}$ term). (c) Summertime predictability, determined from the full LIM. (d) Change in summertime predictability due to the effect of tropical heating (that is, from the $\mathbf{B}_{\text{H}\Psi}$ term).

In a globally averaged sense, maximum possible $S_{\text{global}}(\tau) = [\Sigma_1^2 / \delta(\tau)]^{1/2}$, where

$$\delta(\tau) = \frac{\langle \epsilon^T \epsilon \rangle}{\text{tr}[\mathbf{C}(0)]}$$

is the normalized globally integrated LIM error variance which saturates relatively early, reaching 0.8 within about 10 days. $S_{\text{global}}(\tau)$ is a maximum at about 18 (14) days during winter (summer), so that in a global sense streamfunction predictability could potentially *increase* until then. Locally, maximum possible skill during winter for $\tau=21$ d is larger than it is for $\tau=14$ d (not shown) in a few regions, most notably in a zone extending from the western U.S. to the central Atlantic. This aspect of the LIM mimics the “return-of-skill” (Anderson and van den Dool 1994) seen in some low-dimensional chaotic systems (e.g., Smith et al 1999), but here it arises partly because the noise saturates well before the time tropical heating needs to produce a predictable linear extratropical response. However, it generally requires a very strong projection on the leading singular vector which occurred in only a few percent of all forecasts. On the other hand, about 40% of all week 3 cross-validated

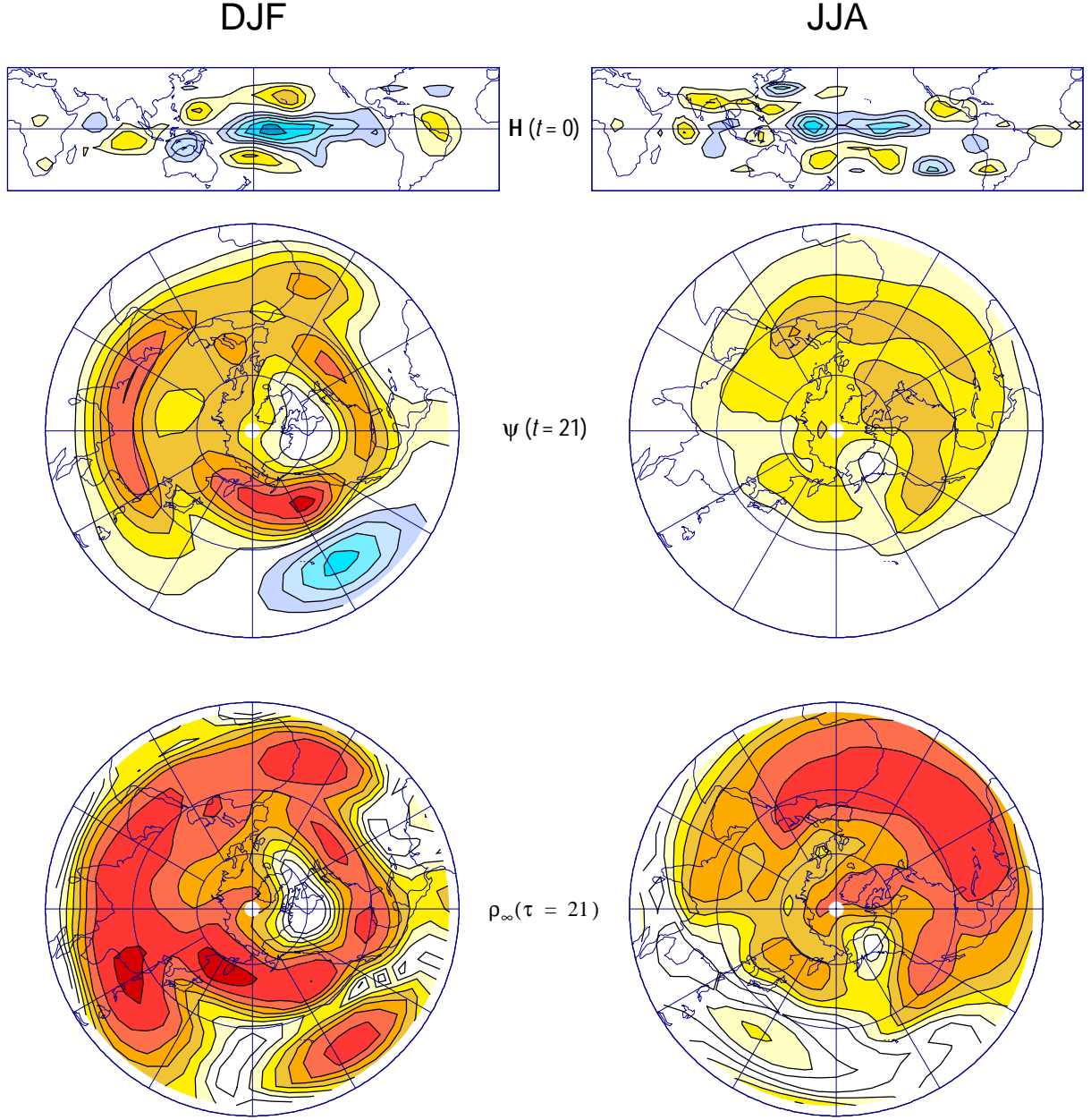


FIG. 10: Initial heating component of the leading singular vector for winter (DJF) and summer (JJA) (top panels), the “final” streamfunction component of that singular vector at day 21 (middle panels), and maps of the associated maximum theoretical predictability (bottom panels). (a) Initial condition of tropical heating during winter. (b) Initial condition of tropical heating during summer. (c) 250 hPa streamfunction 21 days later during winter. (d) 250 hPa streamfunction 21 days later during summer. A heating contour interval of 50 W m^{-2} corresponds to streamfunction contours of $4.9 \times 10^6 \text{ m}^2 \text{ s}^{-1}$. Positive values are denoted by red-yellow shading; negative values are denoted by blue shading. (e and f) Theoretical maximum predictability limit ρ_{∞} for week 3 streamfunction LIM forecasts, for winter (e) and summer (f), at 250 hPa, for cases in which the initial condition is identical to the leading singular vector for $\tau=21 \text{ d}$.

forecasts had higher skill than the week 2 forecast initialized at the same time. This is an expected result of the low signal-to-noise ratio and the small decrease in both expected and actual mean forecast skill from week 2 to week 3 (see Table 3), and thus mostly reflects random variations of the actual noise rather than true return-of-skill.

5.4. State-dependent predictability

Finally, case to case variations of expected forecast skill, associated with predictable variations of signal, are compared to actual skill. First, each initial condition $\mathbf{x}_0(t)$ is used to determine ρ_{∞} from (7) and (9) where $\mathbf{F}(\tau) = \mathbf{G}(\tau)\mathbf{x}_0\mathbf{x}_0^T\mathbf{G}(\tau)^T$. Although significant, the correlation between the daily values of ρ_{∞} and actual daily

forecast skill ρ is not high, however, most likely because noise produces a wide ensemble spread on a daily and even weekly basis.

The effect of noise is reduced by averaging both ρ and ρ_∞ over all the forecasts from a single season, as shown in each panel of Fig. 4. The overall means of these values are included in Tables 3 and 4. Predictability undergoes pronounced interannual variation, due entirely to the interannual variation of the initial conditions and their subsequent linear evolution. Observed and expected skill are not identical, but they now are more strongly correlated, moreso during winter ($r=0.7$) than summer ($r=0.6$). Certainly, the effects of noise are such that we still do not expect this correlation to be 1; that is, even if the LIM were a perfect model the noise PDF would not be well sampled over the course of a single season. Nevertheless, Fig. 4 suggests that much of the interannual variation of both the LIM and MRF98 forecast skill can be understood as interannual variation of linear predictability.

For the LIM, predictability is state-dependent, with higher predictability and higher actual forecast skill associated with singular vector growth. We can demonstrate this by again comparing predicted and actual LIM skill with MRF98 skill, but this time by averaging the skill over cases in which the initial condition projects strongly upon some combination of the three leading (right) singular vectors. Results are shown in Fig. 11. Skill is binned into eight categories, depending upon the projection of the initial condition upon each of the singular vectors. These projections are determined separately for each season and for lead times of both 14 and 21 d. A “high” projection is defined as having amplitude (of either sign) greater than 1σ , as determined from the variance of the projection of all data upon each singular vector. Note that the categories are not equally populated; the largest single category has small projection upon all three leading singular vectors, but generally there is high projection upon at least one and occasionally upon all three.

As expected, predicted skill generally increases as the initial condition projects more strongly on growing structures. The actual LIM skill varies in a similar manner, perhaps most notably for Week 3 forecasts during winter. This relationship can be used to identify *a priori* those LIM forecasts which have high potential for skill and those with high potential for failure; that is, we can “forecast the forecast skill” to some extent. The similar (if weaker) relationship between MRF98 forecast skill and LIM predicted skill suggests that the LIM could be used to identify skillful forecasts of this (and perhaps other) version of the MRF in advance, even when LIM skill is less than MRF skill. Deficiencies in the LIM predicted-MRF98 actual skill relationship may also be useful in helping to identify those forecast scenarios in which model error is, or is not, playing a pronounced role. For example, Fig. 11 might be interpreted as suggesting that the MRF98 does poorer than expected when the projection on both the first and second singular vec-

tors is high. Conversely, during summer MRF98 skill is relatively higher and compares well with the LIM, at both weeks 2 and 3, when the initial conditions strongly project on the second singular vector. At 250 mb, this pattern (not shown) represents an amplification and southeastward extension (weakening and northwestward contraction) of the subtropical Pacific trough centered just north of Hawaii plus a weaker wavetrain extending to the United States, primarily forced by anomalous heating (cooling) in the east Pacific. MRF98 skill in these cases may be less impacted by model error, since the LIM and MRF98 tropical heating skill in the east Pacific are comparable and since the MRF98 systematic error (Fig. 6) is weaker in this region.

Skill of the LIM and MRF98 can be comparable even when the initial condition has small amplitude in the tropics (e.g., wintertime Week 2 singular vector 3), a reminder that predictability in the LIM also results from entirely extratropical dynamics (WNS). Finally, note that the MRF98 has generally higher skill than the LIM when the projection on all three singular vectors is low, suggesting additional sources of skill beyond those estimated by the LIM. Whether this skill is due to predictable nonlinearity captured by the MRF, or simply additional growing (linear) singular vectors not resolved with the limited truncation employed in this version of the LIM, is not known.

6. Concluding remarks

Predictability has often been estimated using complicated nonlinear numerical models and treating them as “perfect” models (e.g., Lorenz 1982; Schubert et al. 1992). These models produce more realistic variability than lower-order “toy” models of chaos, but a complete determination of error growth (for example, computing all the Lyapunov exponents) is not practical. Also, as is well known, such models can have significant deficiencies which degrade their realism. In particular, estimating predictability on subseasonal timescales (indeed, even making forecasts on these timescales) may be seriously compromised both because of the low spread of ensembles (e.g., Whitaker and Lough 1998) and because of the generally poor forecasts of tropical heating made by these models (Ferranti et al. 1990; Slingo et al. 1996; Hendon et al. 2000).

In this paper, we followed a different approach based on the idea that some highly nonlinear systems can be approximated as stochastically forced linear systems. This multivariate red noise approximation proved to be surprisingly accurate on the intermediate timescales of interest here, resulting in empirical linear operators for winter and summer which were of relatively low dimension, yet produced forecasts which were not only about as skillful as the higher-order fully nonlinear numerical models, but also had very similar geographical, temporal, and seasonal variations of forecast skill. These variations of skill were understood as resulting from

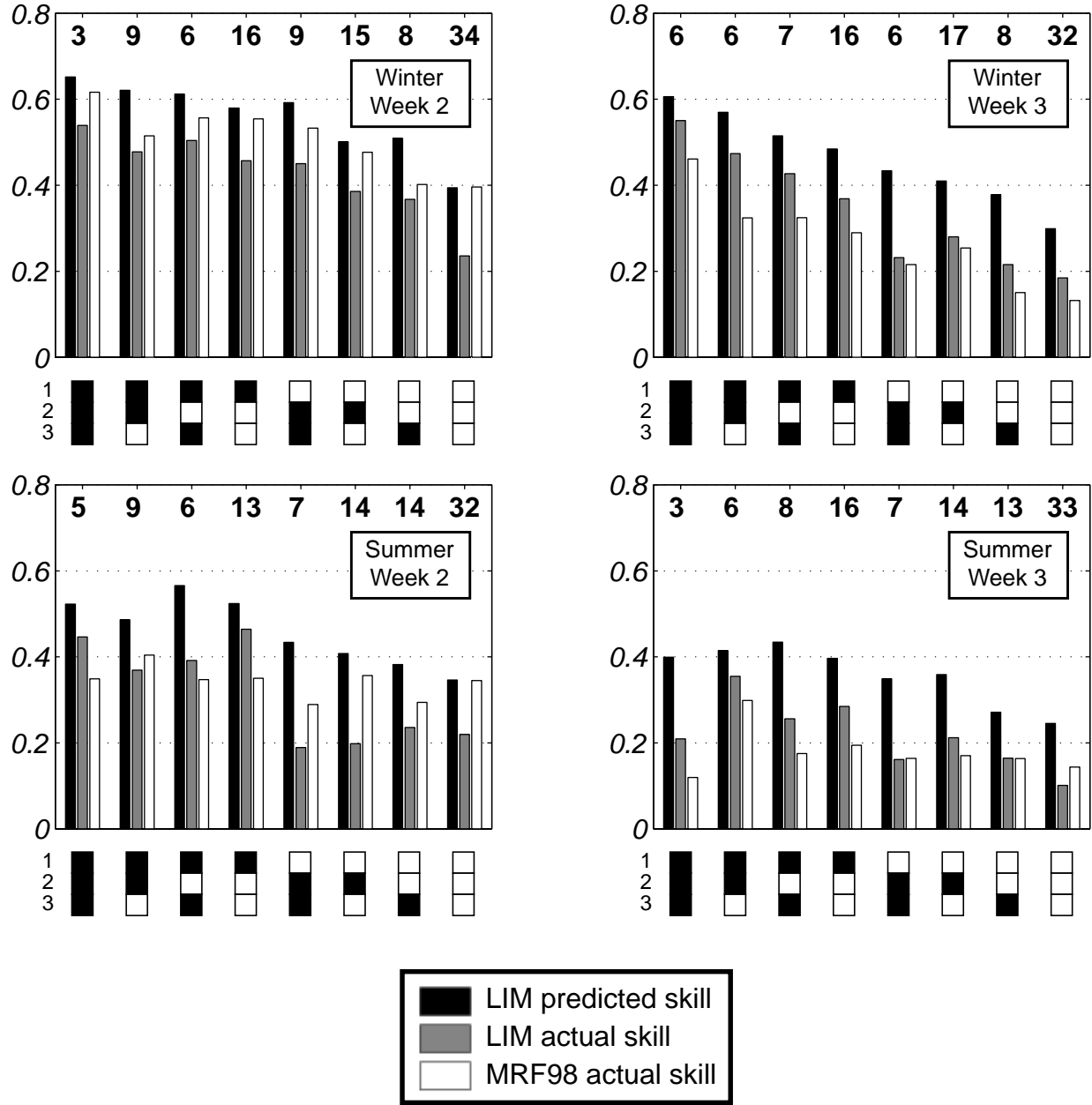


FIG. 11. Mean pattern correlation stratified by projection of the initial condition on the three leading singular vectors, determined separately for both winter and summer and for $\tau = 14$ d and 21 d. Categories are indicated by the filled boxes; for example, an initial condition with high projection on right singular vectors 1 and 3 but low projection on right singular vector 2 is indicated by a black box for 1, a white box for 2, and a black box for 3. The percentage of cases which fall in each category is indicated by the number above the bars, rounded to the nearest integer.

variations of predictability assessed from the relative importance of the deterministic, linear dynamics and the unpredictable noise.

The fact that actual model skill is not as high as the theoretical predictability limit could certainly be expected from limitations in applying the LIM technique to observed data which is neither completely sampled nor entirely accurate. Of course, if predictable nonlinearity exists in the real atmosphere on subseasonal time scales then more sophisticated nonlinear

techniques may ultimately surpass this analysis. On the other hand, we could also add greater sophistication to our model, while retaining the linear nature of the predictable dynamics, by allowing F_s to be red or even to remain state-dependent. For example, in the Tropics F_s could depend upon the phase of the MJO (e.g., Straub and Kiladis 2002). Such multiplicative noise (e.g., Penland 1996; Buizza et al. 1999; Sardeshmukh et al. 2001), although difficult to determine empirically, would make the ensemble spread state-dependent and would thus

generally alter $S(\tau)$. We suspect that this effect is relatively small, however, since not only is ρ_∞ a useful predictor of ρ but also the LIM successfully reproduces the simultaneous and lagged covariability statistics of the real system (Newman et al 2003).

Most of the predictability of the LIM is associated with the growing singular vectors of the LIM's propagator matrix $\mathbf{G}(\tau)$. When the initial projection on these singular vectors is high, not only the potential but also the actual forecast skill is considerably higher than at other times. The version of the NCEP model examined here also appears to have similar variation between relatively high and low forecast skill cases. However, its longer range skill is generally less than that of the LIM, at least partly because of its lesser ability to forecast tropical heating variations in regions where the evolution of the heating component of the growing singular vectors is important to the extratropical response. Our results thus suggest both that there is considerable predictability of Week 2 and Week 3 averages in some regions of the Northern Hemisphere, much of which may be identified *a priori*, and that this skill may not currently be fully realized in some forecasting systems because of their error in predicting tropical heating variations in some sensitive areas. Moreover, since each LIM forecast ensemble takes a few seconds of CPU, extremely cheap and potentially useful Week 3 forecasts are available now.

Acknowledgements. Conversations with many co-workers here at CDC proved invaluable to the development of this project. We would particularly like to thank Cécile Penland and Gil Compo, as well as two anonymous reviewers, for their comments. This work was partially funded by the Predictability DRI of the Office of Naval Research, grant N00014-99-1-0021.

REFERENCES

- Anderson, J., and H. M. van den Dool, 1994: Skill and return of skill in dynamic extended-range forecasts. *Mon. Wea. Rev.*, **122**, 507-516.
- Anderson, J., H. M. van den Dool, A. Barnston, W. Chen, W. Stern, and J. Ploshay, 1999: Present-day capabilities of numerical and statistical models for atmospheric extratropical seasonal simulation and prediction. *Bull. Amer. Meteor. Soc.*, **80**, 1349-1361.
- Brankovic, C., and T. N. Palmer, 1997: Atmospheric seasonal predictability and estimates of ensemble size. *Mon. Wea. Rev.*, **125**, 859-874.
- Branstator, G., A. Mai, and D. Baumhefner, 1993: Identification of highly predictable flow elements for spatial filtering of Medium- and Extended-Range numerical forecasts. *Mon. Wea. Rev.*, **121**, 1786-1802.
- Buizza, R., M. Miller, and T. N. Palmer, 1999: Stochastic representation of model uncertainties in the ECMWF Ensemble Prediction System. *Quart. J. Roy. Meteor. Soc.*, **125**, 2887-2908.
- Cai, M., J. S. Whitaker, R. M. Dole, and K. L. Paine, 1996: Dynamics of systematic errors in the NMC medium range forecast model. *Mon. Wea. Rev.*, **124**, 265-276.
- Caplan, P., J. Derber, W. Gemill, S.-Y. Hong, H.-L. Pan, and D. Parrish, 1997: Changes to the 1995 NCEP Operational Medium-Range Forecast Model Analysis-Forecast System. *Weather and Forecasting*, **12**, 581-594.
- Chen, W. Y., and H. van den Dool, 1999: Significant change of extratropical natural variability and potential predictability associated with the El Niño/Southern Oscillation. *Tellus*, **51A**, 790-802.
- Chervin, R. M., 1986: Interannual variability and seasonal climate predictability. *J. Atmos. Sci.*, **43**, 233-251.
- Corti, S., and T. N. Palmer, 1997: Sensitivity analysis of atmospheric low-frequency variability. *Q. J. R. Meteorol. Soc.*, **123**, 2425-2447.
- DelSole, T., and A. Y. Hou, 1999: Empirical stochastic models for the dominant climate statistics of a general circulation model. *J. Atmos. Sci.*, **56**, 3436-3456.
- Farrell, B. F., 1988: Optimal excitation of neutral Rossby waves. *J. Atmos. Sci.*, **45**, 163-172.
- Fels, S. B. and M. D. Schwarzkopf, 1975: The simplified exchange approximation: A new method for radiative transfer calculations. *J. Atmos. Sci.*, **32**, 1475-1488.
- Ferranti, L., T. N. Palmer, F. Molteni, and E. Klinker, 1990: Tropical-extratropical interaction associated with the 30-60 day oscillation and its impact on medium and extended range prediction. *J. Atmos. Sci.*, **47**, 2177-2199.
- Hasselmann, K., 1976: Stochastic climate models. Part I. Theory. *Tellus*, **28**, 474-485.
- Hendon, H. H., B. Liebmann, M. Newman, and J. D. Glick, 2000: Medium-range forecast errors associated with active episodes of the Madden-Julian oscillation. *Mon. Wea. Rev.*, **128**, 69-86.
- Kalnay, E., and co-authors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437-472.
- Kleeman, R., 2002: Measuring dynamical prediction utility using relative entropy. *J. Atmos. Sci.*, **59**, 2057-2072.
- Klinker, E., 1990: Investigation of systematic errors by relaxation experiments. *Quart. J. Roy. Meteor. Soc.*, **116**, 573-594.
- Kumar, A., M. Hoerling, M. Ji, A. Leetmaa, and P. D. Sardeshmukh, 1996: Assessing a GCM's suitability for making seasonal predictions. *J. Climate*, **9**, 115-129.
- Lacarra, J.-F. and O. Talagrand, 1988: Short range evolution of small perturbations in a barotropic model. *Tellus*, **40A**, 81-95.
- Lacis, A., and J. E. Hansen, 1974: A parameterization for the absorption of solar radiation in the earth's atmosphere. *J. Atmos. Sci.*, **31**, 118-133.
- Leith, C., 1971: Atmospheric predictability and two-dimensional turbulence. *J. Atmos. Sci.*, **28**, 145-161.
- Lorenz, E. N., 1982: Atmospheric predictability experiments with a large numerical model. *Tellus*, **34**, 505-513.
- Madden, R. A., 1976: Estimates of natural variability of time-averaged sea level pressure. *Mon. Wea. Rev.*, **104**, 942-952.
- Newman, M., and P. D. Sardeshmukh, 1998: The impact of the annual cycle on the North Pacific/North American response to remote low-frequency forcing. *J. Atmos. Sci.*, **55**, 1336-1353.
- Newman, M., M. A. Alexander, C. R. Winkler, J. D. Scott, and J. J. Barsugli, 2000: A linear diagnosis of the coupled extratropical ocean atmosphere system in the GFDL GCM. *Atmos. Sci. Lett.*, **1**, 14-25.
- Newman, M., P. D. Sardeshmukh, and C. R. Winkler, 2002: A linear model of wintertime low-frequency variability. Part II: Diagnosis of covariance statistics. To be submitted to *J. Atmos. Sci.*
- Pan, H.-L., and W.-S. Wu, 1994: Implementing a mass-flux convective parameterization package for the NMC Medium Range Forecast Model. Preprints, 10th Conf. on Numerical Weather Prediction, Portland, OR, Amer. Meteor. Soc., 9698.
- Papanicolaou, G. and W. Kohler, 1974: Asymptotic theory of mixing stochastic ordinary differential equations. *Commun. Pure Appl. Math.*, **27**, 641-668.
- Penland, C., 1989: Random forcing and forecasting using principal oscillation pattern analysis. *Mon. Wea. Rev.*, **117**, 2165-2185.
- Penland, C., 1996: A stochastic model of Indo-Pacific sea surface temperature anomalies. *Physica D*, **98**, 534-558.
- Penland, C., and M. Ghil, 1993: Forecasting Northern Hemisphere 700-mb geopotential heights using principal oscillation patterns. *Mon. Wea. Rev.*, **121**, 2355-2372.
- Penland, C., and P. D. Sardeshmukh, 1995: The optimal growth of tropical sea surface temperature anomalies. *J. Climate*, **8**, 1999-2024.

- Qin, J. and W. A. Robinson, 1995: The impact of tropical forcing on extratropical predictability in a simple global model. *J. Atmos. Sci.*, **52**, 3895-3910.
- Rowell, D. P., 1998: Assessing potential seasonal predictability with an ensemble of multidecadal GCM simulations. *J. Climate*, **11**, 109-120.
- Sardeshmukh, P. D., and B. J. Hoskins, 1984: Spatial smoothing on the sphere. *Mon. Wea. Rev.*, **112**, 2524-2529.
- Sardeshmukh, P. D., and B. J. Hoskins, 1985: Vorticity balances in the tropics during the 1982-1983 El Niño-Southern Oscillation event. *Q. J. R. Meteor. Soc.*, **111**, 261-278.
- Sardeshmukh, P. D., and B. J. Hoskins, 1988: The generation of global rotational flow by steady idealized divergence. *J. Atmos. Sci.*, **45**, 1228-1251.
- Sardeshmukh, P. D., 1993: The baroclinic chi problem and its application to the diagnosis of atmospheric heating rates. *J. Atmos. Sci.*, **50**, 1099-1112.
- Sardeshmukh, P. D., M. Newman, and M. D. Borges, 1997: Free barotropic Rossby wave dynamics of the wintertime low-frequency flow. *J. Atmos. Sci.*, **54**, 5-23.
- Sardeshmukh, P. D., M. Newman, and C. R. Winkler, 1999: Dynamically consistent estimates of diabatic heating. Proceedings, *24th Annual Climate Diagnostics and Prediction Workshop*, Tucson, AZ, 172-175.
- Sardeshmukh, P. D., G. P. Compo, and C. Penland, 2000: Changes of probability associated with El Niño. *J. Climate*, **13**, 4268-4286.
- Sardeshmukh, P. D., C. Penland, and M. Newman, 2001: Rossby waves in a fluctuating medium. In *Stochastic Climate Models*, ed. P. Imkeller and J.-S. von Storch, *Progress in Probability*, Birkhauser, Basel, pp. 369-384.
- Schneider, T. and S. M. Griffies, 1999: A conceptual framework for predictability studies. *J. Climate*, **12**, 3133-3155.
- Schubert, S., M. J. Suarez, and J.-K. Schemm, 1992: Persistence and predictability in a perfect model. *J. Atmos. Sci.*, **49**, 256-269.
- Shukla, J. and D. S. Gutzler, 1983: Interannual variability and predictability of 500 mb geopotential heights over the Northern Hemisphere. *Mon. Wea. Rev.*, **111**, 1273-1279.
- Slingo, J. M., and Coauthors, 1996: Intraseasonal oscillations in 15 atmospheric general circulation models: Results from an AMIP diagnostic subproject. *Climate Dyn.*, **12**, 325-357.
- Smith, L. A., C. Ziehmann, and K. Fraedrich, 1999: Uncertainty dynamics and predictability in chaotic systems, *Q. J. R. Meteorol. Soc.*, **125**, 2855-2886.
- Straub, K. H., and G. N. Kiladis, 2002: Observations of a convectively coupled Kelvin Wave in the Eastern Pacific ITCZ. *J. Atmos. Sci.*, **59**, 30-53.
- Tiedtke, M., 1983: The sensitivity of the time-mean large-scale flow to cumulus convection in the ECMWF model. Workshop on convection in large-scale models, ECMWF, 297-316.
- Troen, I. B., and L. Mahrt, 1986: A simple model of the atmospheric boundary layer: Sensitivity to surface evaporation. *Bound.-Layer Meteor.*, **37**, 129-148.
- van den Dool, H. M., and S. Saha, 1990: Frequency dependence in forecast skill. *Mon. Wea. Rev.*, **118**, 128-137.
- Whitaker, J. S., and A. F. Lough, 1998: The relationship between ensemble spread and ensemble mean skill. *Mon. Wea. Rev.*, **126**, 3292-3302.
- Winkler, C. R., M. Newman, and P. D. Sardeshmukh, 2001: A linear model of wintertime low-frequency variability. Part I: Formulation and forecast skill. *J. Climate*, **14**, 4474-4494.
- Wu, W., M. Iredell, S. Saha and P. Caplan, 1997: Changes to the 1997 NCEP Operational MRF Model Analysis/Forecast System. NCEP Technical Procedures Bulletin 443, available from <http://www.nws.noaa.gov/om/tpb/indexb.htm>, or the National Weather Service, Office of Meteorology, Programs and Plans Division, Silver Spring, MD, 20910.
- Zhang, C., and H. H. Hendon, 1997: Propagating and standing components of the intraseasonal oscillation in tropical convection. *J. Atmos. Sci.*, **54**, 741-752.
- Zwiers, F. W., 1987: A potential predictability study conducted with an atmospheric general circulation model. *Mon. Wea. Rev.*, **115**, 2957-2974.